Woman C.P.A.

Volume 44 | Issue 3 Article 7

7-1982

Electronic Data Processing: Perspective On Distributed Data Base Systems

Elise G. Jancura

Alfred R. Michenzi

Advanced systems have been characterized as those systems possessing one or more of the following characteristics:[1]

Data communications
Data integration
Automatic transaction initiation
Unconventional or temporary audit trail

The introduction of these characteristics into the data processing environment is the natural result of the application of computers to a variety of accounting and management information applications and the attempt to provide greater access to computer processing to a variety of potential users in each organization. All of these characteristics are, or can be, present in distributed data base systems.

Integrated Data Files and Data Bases

One of the strong trends in advanced systems has been the integration of data from multiple related applications. Increased data integration can provide both operational and economic advantage. Data integration can minimize the data redundancy which occurs when each application creates and updates individual application-oriented files. Elimination of redundancy or duplication of data elements within application files can promote more efficient use of physical storage facilities, and eliminate some operational procedures such as repetitive file sorting. More importantly, it can promote greater accuracy in the data by eliminating the inconsistencies introduced when lags occur between the updates of redundant data or when all updates of the redundant data do not produce the same results.

Data integration also provides economies in the development and maintenance of application software. File definition functions are removed from the individual application programs. This makes original development and modification of these applications more simple, while at the same time changes to the data base (physical organization or new elements) will not require program changes to all application programs using the integrated base.

Usually data integration is accomplished through a formal definition of individual data elements or items and the relationship of the data elements to each other and the various application programs processing them. When the relationships are formalized, the integrated data is a data base. One definition of a data base is "a collection of stored operational data used by the application systems of some particular enterprise."[2] Another definition is: "A data base is a collection of files which can be accessed via the use of a data base management system and among which relationships are defined which play a role in those applications which use the data contained in the files."[3]

A data base management system (DBMS) can be defined as a software system intended to manage and maintain data in a nonredundant structure for the purpose of being processed by multiple applications. The DBMS organizes data elements in some predefined structure, and retains relationships between different data elements within the data base. Among other characteristics, a DBMS may make it possible for non-programming personnel to utilize the data located in the data base with reduced effort through the use of query languages designed specifically to meet the needs of users of the system. It is the DBMS which translates the data requirements of application programs and relates then to the processing functions of data management such as physical data storage and organization, logical organization, and data retrieval and maintenance.

Implications for Control in Data Base Systems

With data integration, changes occur not only in the physical and logical relationships of the data, but also in the responsibilities and relationship of the users and the data processing personnel. As a data base is developed and a data base management system (DBMS) implemented, the responsibility for the physical storage, the logical organization of data, and the subsequent ability to access and process an organization's data becomes increasingly centralized in the programs of the DBMS. Application programs no longer access data or manipulate the data files directly. The DBMS acts as an interface to all the programs using the data base. It is the DBMS which determines which data is to be accessed, which performs all physical input/output operations, and maintains the organization of the data base. Thus, many controls over access to data shift from the user and/or individual applications. This makes controls over access more critical. The concern encompasses both the question of which users may access individual data elements and which program or processing functions may be applied to individual data elements.

The concentration of control in the DBMS over the data base provides a potential for strengthening the procedures employed to insure completeness and accuracy of data as well as controlled access. At the same time, the increased concentration does pose some difficulties. First, many of the controls are programmed as part of the software. This introduces an increased technical sophistication that must be

Electronic Data Processing

Perspective On Distributed Data Base Systems

Editor:
Elise G. Jancura, CPA, CISA, Ph.D
The Cleveland State University
Cleveland, Ohio 4415

By Elise G. Jancura and Alfred R. Michenzi

mastered if the level of control is to be properly evaluated and used. Secondly, there may be some tendency by users to abdicate some responsibilities to the DBMS. Users must continue to share in the process of checking on the accuracy and completeness of individual data elements by retaining their strong responsibilities for input and output controls.

The concentration of the many control functions into the DBMS itself can weaken the separation of responsibilities. Thus, while the data management function must be centralized (usually in the data base administrator function) to achieve the degree of coordination necessary to make the data base function, it is essential that individual user departments be given the authority and responsibility for the definition and content of the individual data elements, the definition of what access may be made of those elements, and the continued verification of the accuracy of "their" data elements. Security and accuracy of the data must be accomplished by a combination of procedures in both the data processing department and the user department. A precise definition of the responsibilities at both levels is essential.

Documentation in a data base becomes a critical need. The integration of data means that the formal definition of logical relationships among the data are essential to maintain proper relationships among data elements and to provide a reliable mechanism to access specified data elements. The documentation must also carefully define the physical layout of the data, and the procedures for user and program identification.

Distributed Data Bases

A distributed data base system (DDBS) exists when the data elements stored at multiple locations are interrelated, or when a process (program execution) at one location requires access to data stored at another location. Thus, a DDBS always exists within an information networking environment.

> The network provides the underlying configuration of computer systems and communication facilities within which data is stored, DBMS's operate, and users access data. A node in the network consists of computer processing facilities (ranging from a large multiprocessor computer to an intelligent terminal) and an associated operating system sufficient for executing user and DBMS processes (programs, queries, etc.). In addition, data and its definition may be stored at a node. The precise structure of a node is an architectural design choice independent of the manner in which the node is connected to other nodes and the extent of geographical separation.
>
> The communications facility is the collection of processes and physical facilities which interconnect the nodes. The communications facility includes knowledge of the physical location of each node, the physical path connections between the nodes, and the protocols to be used in sending messages between nodes. Processes in the communications facility will accept a message from one node and deliver it to another node or broadcast it to some or all other nodes. Two nodes may be connected directly or indirectly through other nodes. A network access process (NAP) exists at every node as the interface between processes at the node and the communications facility. The NAP is that portion of the communications facility which executes on the processing facilities of a node.[4]

A DDBS provides the potential advantage of the efficiencies of data integration, while at the same time providing for greater flexibility in the configuration of the system and optimal distribution of "processing power" through the ability to decentralize processing facilities.

A major concern of the user of data has been a need for easy, fast access to data. At the same time as organizations convert increasing proportions of their financial and operating data to computerized systems, the number, the functional variety, and the geographic dispersion of the "end users" of computer data make the concept of distributed processing more attractive. The options available in distributed data processing allows users to play more active roles in data processing and on occasion to actually control their own computing resource without sacrificing the benefits of integrated data files or other centralized processing activities. Thus in some aspects, distributed processing allows the data processing function to more closely approximate the organization of management.

Distributed systems sometimes make it possible to employ more specialized equipment to specific tasks within the processing spectrum and in some instances are the only logical alternative when a single centralized system doesn't have the capacity for the total job. Also growth can be more easily accommodated with the smaller increments possible in a distributed system.

The addition of the data communications element does introduce additional implications for control. The procedure for identification and authorization of users, maintenance of logical and physical relationships, and all other techniques for controlling access to the system must be expanded to all the locations in the system. Frequently these controls will be programmed producing even greater dependence on the system and system software for adequate control. In addition to checking data accuracy and completeness, controls must be introduced to verify transmission activities and the operations at each node.

Design Alternatives for Distributed Data Bases

There are several alternate approaches to the organization of a DDBS. One method of classifying a system is by the way in which the data is organized and distributed. One approach is centralized data storage with distributed processing facilities. Another approach is that in which the data itself is also distributed.

Partitioned data bases exist when a single copy of the data base is separated into segments and the segments are stored at different nodes or processing facility. These partitions usually form a "logical" database because of the interrelationship between the segments. Usually the segments or partitions are formed or grouped in response to some natural distribution of access requirements. This usually helps minimize transmission costs between the nodes.

Replicated data bases occur when multiple copies of a database or its pieces are stored at multiple nodes. This redundancy may be tolerated because it facilitates increased and more efficient accessibility, provides readily available backup, and provides decreased communication time. However, the redundancy causes some additional cost and complexity in updating the file and requires more storage capacity. A frequent approach of the replicated data base occurs when the central location contains a full master file and the remote nodes each contain a copy of a segment of the master

file. Each local data base is created by copying data from one area of the central data base. The local data bases are used for local processing, but local processing does not directly affect the central data base and does not directly update the local node. Instead, the central data base is updated centrallly on a periodic basis from the accumulated transactions of all the locals and then used to make "new" copies of the local data bases.

DDBS may also be categorized by the distribution of the processing functions. Horizontal distribution of processing functions occurs when the components which are interconnected are logically equal although they may be physically diverse with different capacity and power. Frequently the total workload is distributed because the processing components exchange jobs and/or data cooperatively. While each node usually handles "local" jobs or transactions, the goal of "load leveling" may promote exchanges of processing when overloads occur at any given node.

In a hierarchical or vertical distribution, each component or mode is controlled to some degree by the higher-level nodes and the processing load is distributed up and down the hierarchy. Usually "high volume-fast response" items are located as "low" in the hierarchy as possible, while lower volume, "slower-response" functions would be moved "up" the hierarchy.

The variations in data distribution (partition or replication) and in processing distribution (horizontal or hierarchical) can be combined in a variety of ways. Further, the distributed data base may operate in either batch or in real-time mode (sometimes referred to as asynchronous or synchronous mode, respectively).

Audit & Control Concerns in Distributed Data Bases

Distributed data base systems can have a significant impact on internal control features and the question arises whether they may require the development and use of new or additional auditing techniques in the performance of both compliance and substantive testing.

Because control is more heavily invested in the EDP system in a distributed data base, the nature of the system controls become more critical. The AICPA Task Force on Auditing Advanced EDP Systems listed the following internal control requirements in advanced systems:

> Adequate control features and procedures must be developed for communication-based networks and distributed systems in which accounting information can be accessed or changed from remote locations.
>
> Authorization systems are required to control access to and processing of accounting information and to maintain a separation of employee functions.
>
> Programmed system controls must be provided since a manual review of input by employees will no longer be applicable when accounting transactions are generated and processed automatically by the system.
>
> Provisions for tracing the historical flow of accounting transactions should be provided in systems having accounting significance.
>
> Provisions should be made for timely and economical reconstruction of accounting information in the event of its destruction.
>
> Management, auditors, and others should be provided with feedback on the performance and integrity of advanced EDP systems.[5]

The Task Force also suggests that the areas of difference between conventional and advanced systems, from an auditing point of view, include:

- Complexity
- Nature of evidential matter
- Relationship between accounting control and evidential matter
- Need for audit control
- Audit trail considerations
- Techniques required for access to information
- Timing of audit procedures[6]

In a DDBS the auditor will likely find it necessary to use the system itself to generate and/or collect necessary evidential matter. This will require a greater technical familiarity with the system. Further, the auditor will have to be concerned with assuring proper audit control. Where the only audit evidence available is machine-sensible data, the auditor may not have an alternative to reliance on the accounting controls in the system and the corresponding requirement for increased compliance testing.

The timing of audit tests may also change. In batch systems, collection of historical data may be adequate for audit purposes. But data base systems rarely remain static and when the system is operating in an interactive mode, the auditor may be forced to shift the timing of audit tests. Thus, consideration of "concurrent auditing" may be appropriate where the audit tests occur at a point immediately following or concurrent with the occurrence of the transaction. Another alternative is the controlled use by the auditor of the systems log to capture audit data.

The Task Force has suggested a number of audit techniques which might be useful in advanced systems. The techniques are summarized in Table 1.1.

Impact on Audit Techniques

Audit objectives are not changed by the introduction of distributed processing, data communications, or integrated data base systems. The auditor's responsibilities include a proper study and evaluation of the system of internal control, execution of appropriate compliance tests if reliance is to be placed on internal control, and execution of appropriate substantive tests of the financial data.

Whenever possible, the auditor should seek independent evidence of processing controls or data accuracy. Techniques such as confirmations and inventory counts are still most appropriate. However, systems containing the features discussed earlier tend to reduce the amount of independent source data and manual intervention.

Auditors can continue to use computer-assisted techniques that operate on historical data and compliance testing techniques which use simulated or test data. In addition, however, the ability to use live data on a real-time basis in both the compliance and substantive tests is greatly facilitated by the availability of the system characteristics discussed. The ability to use live data on a real-time basis requires that a prewritten audit program be executed as the data is being processed. This audit program may be incorporated into the operating system or the application program. Successful use of such a technique, however, requires that sufficient controls exist to prevent modification or circumvention of the audit programming in place. This approach requires considerable advance planning and usually can be successfully implemented if adequate attention is devoted to control and audit requirements at the systems design stage. Ω

NOTES

[1]American Institute of Certified Public Accountants: Audited Advanced EDP Systems

TABLE 1.1
Techniques Matrix

Technique	Capability Supplied by	Used by	Data Used	Purpose	Advantages	Disadvantages
Transaction tagging	Vendor or application system designer	Auditors and managers	Live accounting	Compliance and substantive test	Full range of selectivity	Adds to overhead of system, special programming
Real time notification	Systems programmer or vendor	Auditors and managers	Live accounting and system	Compliance test and control	Control and timeliness	Cost
Audit log	System designer	Auditors and control personnel	Live accounting and system	Compliance and substantive test	Specified transactions logged for audit review	Cost
Monitoring	Vendor	Auditors and managers	Live system	Review actual system activity	Shows what has happened	Requires technical knowledge to interpret
Audit language and programs	Vendor and system designer, software house, manufacturer or audit firm	Auditors and managers	Historical and live	Compliance and substantive test. Perform wide variety of audit tests	Retrieves data for audit purposes. Relatively easy to use, not expensive	Requires some programming knowledge by auditor. Presently limited to types of files that can be accessed.
Simulation	Auditors, internal and external with program copy	Auditors	Historical	Determine accuracy of data processed	Permits comparison with real processing	Extensive use can be large consumer of machine resources
Extended records	Design of client applications	Auditors and managers	Historical	Provide complete trail for audit and management purposes	Provides complete account history	Very costly use of machine resources at present
Integrated test facility	Auditors, mostly internal	Auditors	Dummy	Compliance test	Relatively inexpensive	Must be "backed out" very carefully
Program analysis techniques	Special software, contractor or vendor	Auditors and programmers	Usually dummy	Authentication of program operation. Check of key points in program execution	Gives better understanding of application; gives assurance controls are functioning	Needs auditor knowledge of programming, may be expensive; useful only in certain circumstances.

Taken from: American Institute of Certified Public Accountants: Auditing Advanced EDP Systems Task Force, *Management, Control and Audit of Advanced EDP Systems,* (AICPA, 1977)

Task Force, *Management, Control and Audit of Advanced EDP Systems,* (AICPA, 1977), p.5.

[2]R.W. Engles, "A Tutorial on Data Base Organization," *Annual Review in Automatic Programming,* Vol. 7, Part I, Halpern and McGee, Ed., (Pergamon Press, July, 1972), p.32.

[3]J.H. Balvert, K. Gerritse, H. Roos, K. VanTilburg, J. DeVas, P.C. Warners, *Data Base and the Accountant,* (Sampson Uitgenerij Alphen Aan den Rijor, 1977), p. 21.

[4]CODASYL System Committee, "Distributed Data Base Technology-An Interim Report," *Proceedings of the National Computer Conference,* (1978) p.910.

[5]*Management, Control, and Audit of Advanced EDP Systems,* pp.1-2.

[6]*Management, Control, and Audit of Advanced EDP Systems,* p.15.

Elise G. Jancura, *CPA, CISA, Ph.D., is chairperson of the Accounting and Business Law Departments of Cleveland State University. She has served on the Computer Services Executive Committee with AICPA and as chairperson of the Computer Education Committee. She is a member of the Ohio Society of CPAs, the Assoc. for Computing Machinery, and is the immediate past president of the Cleveland Chapter of American Society of Women Accountants.*

Alfred R. Michenzi, *CPA, Ph.D., is associate professor of accounting at Cleveland State University. He is a member of AICPA, the Ohio Society of CPAs, and has research interests in the areas of auditing, computer auditing, accounting and information systems, and application of quantitative techniques in auditing. He has been published in various professional journals.*